


RESEARCH ARTICLE

Open Access



# Hybridization and introgression between *Helicoverpa armigera* and *H. zea*: an adaptational bridge

Erick M. G. Cordeiro<sup>1†</sup>, Laura M. Pantoja-Gomez<sup>1†</sup>, Julia B. de Paiva<sup>1</sup>, Antônio R. B. Nascimento<sup>1,2</sup>, Celso Omoto<sup>1</sup>, Andrew P. Michel<sup>2</sup> and Alberto S. Correa<sup>1\*</sup> 

## Abstract

**Background:** Invasion of organisms into new ecosystems is increasingly common, due to the global trade in commodities. One of the most complex post-invasion scenarios occurs when an invasive species is related to a native pest, and even more so when they can hybridize and produce fertile progeny. The global pest *Helicoverpa armigera* was first detected in Brazil in 2013 and generated a wave of speculations about the possibility of hybridization with the native sister taxon *Helicoverpa zea*. In the present study, we used genome-wide single nucleotide polymorphisms from field-collected individuals to estimate hybridization between *H. armigera* and *H. zea* in different Brazilian agricultural landscapes.

**Results:** The frequency of hybridization varied from 15 to 30% depending on the statistical analyses. These methods showed more congruence in estimating that hybrids contained approximately 10% mixed ancestry (i.e. introgression) from either species. Hybridization also varied considerably depending on the geographic locations where the sample was collected, forming a 'mosaic' hybrid zone where introgression may be facilitated by environmental and landscape variables. Both landscape composition and bioclimatic variables indicated that maize and soybean cropland are the main factors responsible for high levels of introgression in agricultural landscapes. The impact of multiple *H. armigera* incursions is reflected in the structured and inbred pattern of genetic diversity.

**Conclusions:** Our data showed that the landscape composition and bioclimatic variables influence the introgression rate between *H. armigera* and *H. zea* in agricultural areas. Continuous monitoring of the hybridization process in the field is necessary, since agricultural expansion, climatic fluctuations, changing composition of crop species and varieties, and dynamic planting seasons are some factors in South America that could cause a sudden alteration in the introgression rate between *Helicoverpa* species. Introgression between invasive and native pests can dramatically impact the evolution of host ranges and resistance management.

**Keywords:** Hybrid zone, SNP, Population genomics, Introgression, World pest, Landscape genomics

\* Correspondence: [ascorrea@usp.br](mailto:ascorrea@usp.br)

Submitted to BMC Evolutionary Biology

<sup>†</sup>Erick M. G. Cordeiro and Laura M. Pantoja-Gomez contributed equally to this work.

<sup>1</sup>Department of Entomology and Acarology, University of São Paulo, Luiz de Queiroz College of Agriculture, Piracicaba, São Paulo 13418900, Brazil

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

An invasive pest can cause adverse effects of various degrees of severity, as high adaptation potentials and dispersal can cause dramatic costs to ecosystem services and agricultural production [1–4]. Managing these costs is significantly more difficult in cases where the invasive species is related to a native species and is exacerbated when there is potential for fertile hybridization [5]. The ‘hybrid bridge’ hypothesis provides a mechanism for host shifting and host expansion in herbivorous insect pests and suggests that hybridization events might combine lineage-specific adaptations within a single organism [6]. Interspecific gene flow (introgression) can be uni- or bi-directional and facilitated by the ecological context of the interaction between the two species [6]. Due to potential differences in introgression, the proper diagnosis of hybridization encounters serious difficulties since, at the genomic level, markers must be genome wide (to identify areas of introgression) and distinguish between true species [7, 8]. Without such information, challenges will persist for improving pest management and mitigating the effects of invasive species [8, 9].

The invasive bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) is native to the Old World (Asia, Europe, Africa, and Australasia) and is one of the most important pests worldwide [8]. This insect has an annual impact of billions of dollars, caused by crop damage and the high cost of pest control [5]. For those reasons, *H. armigera* is a threat for crops in the New World and is designated a quarantine pest in many countries, including Brazil. Since the first report from the Americas in 2013 [5], much research has been devoted to understanding its potential for global spread [10]. *Helicoverpa armigera* is now a ‘world mega pest’ because of its rapid evolution of resistance to synthetic insecticides and, more recently, to genetically modified plants containing *Bt* protein [8, 11, 12].

Other species of *Helicoverpa*, such as the corn earworm (*H. zea*), are present in many New World countries. *Helicoverpa zea* is morphologically similar to *H. armigera*, and these two species diverged around 1.5 Mya [13]. Although the evolutionary relationship between *H. armigera* and *H. zea* is not fully understood, the two species appear to be monophyletic with the common ancestor *H. assulta* [14]. *Helicoverpa zea* likely derived from a small population of *H. armigera* that invaded areas of the New World in the past, which may explain the lower destructive capacity of *H. zea* compared to *H. armigera* [13].

Unlike other congeners, *H. zea* and *H. armigera* are highly polyphagous and can produce fertile hybrids [15, 16]. A complex pattern of genetic structure and gene flow exists within *H. armigera* populations across the globe [17–22]. Genetic diversity and structure could be

attributed to interactions between agricultural practices and the life history of the organism. Adding to this complexity are differences in the molecular markers among studies that can include isoenzymes [17], mitochondrial DNA [18], sodium channel sequences [19], and microsatellites [20, 21]; these studies have not found clear, fine-scale genetic structure. Nevertheless, the high gene flow, low genetic differentiation, and large effective population sizes are common occurrences in insect pest moth species, including most *Helicoverpa* populations worldwide [22, 23].

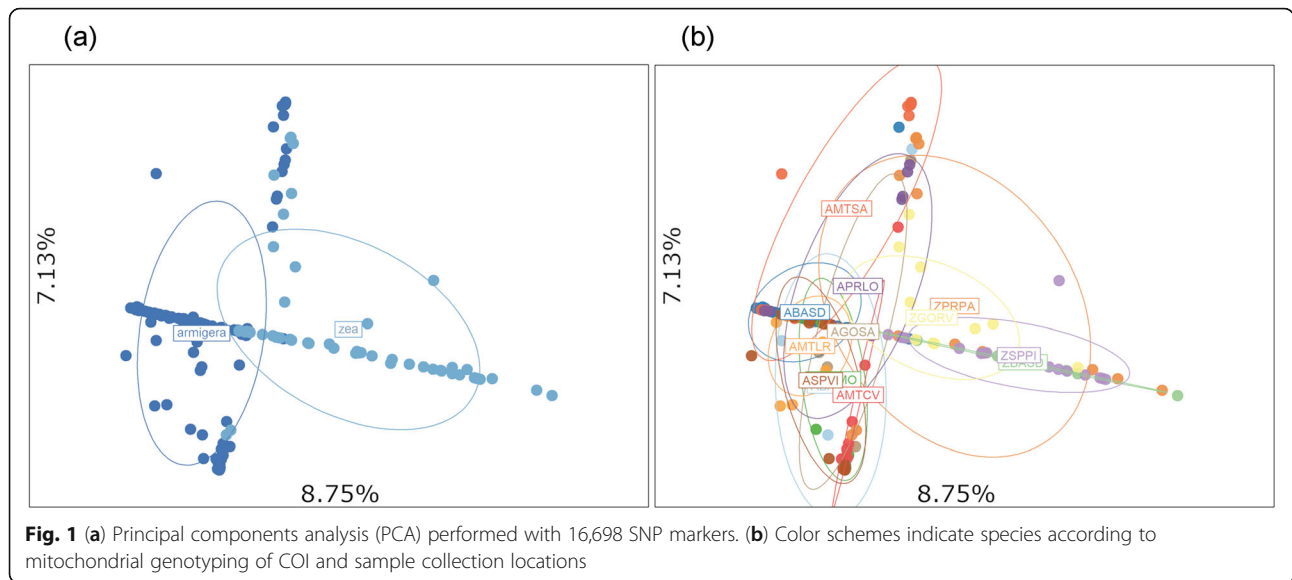
After the South American invasion, both *Helicoverpa* species have coexisted in the complex host compositions across the Brazilian agriculture landscapes. These landscapes generally consist of a large number of crops that form a mosaic with natural areas. More intensively farmed areas such as the Cerrado (central high plateau), are dominated by cotton, soybean, and maize [24]. In Brazil, *H. zea* is a primary pest of maize (monocotyledons), whereas *H. armigera* feeds primarily on soybean and cotton (dicotyledons). Hybridization could result in more intense pressure of caterpillar feeding on soybean, and introgression of *H. zea* genes associated with resistance to pesticides and *Bt* crops into *H. armigera*. The potential for hybridization requires additional validation with more powerful markers providing sufficient resolution to detect introgression.

In this study, we used genome-wide single nucleotide polymorphisms (SNPs) to detect hybrids in the most critical agricultural production areas in Brazil. We also quantified the extent of introgression, which was correlated with landscape and environmental attributes and appeared to facilitate hybridization. In a broader context, this research can improve our understanding of how rapidly changing ecosystems favor evolutionary adaptation through hybridization between native and invasive species.

## Results

### Genetic structure and hybridization

The non-model-based PCA generated two clusters corresponding to mitochondrial identification, using a fragment of the COI region (Fig. 1a). The data showed detectable overlapping between genetic groups, indicating possible hybridization events occurring at a minimum of five locations: AGOSA, APRLO, AMTSA, AMTCV, and ZPRPA (Fig. 1b). Calculations included putative hybrids and pure-bred insects and were based on 16,698 SNP markers. Fixation index ( $F_{ST}$ ) among sampling locations showed a high degree of genetic differentiation, with a mean value of 0.264. Much higher genetic differentiation occurred among *H. armigera* ( $F_{ST} = 0.23$ ) compared to *H. zea* ( $F_{ST} = 0.07$ ) Pairwise  $F_{ST}$

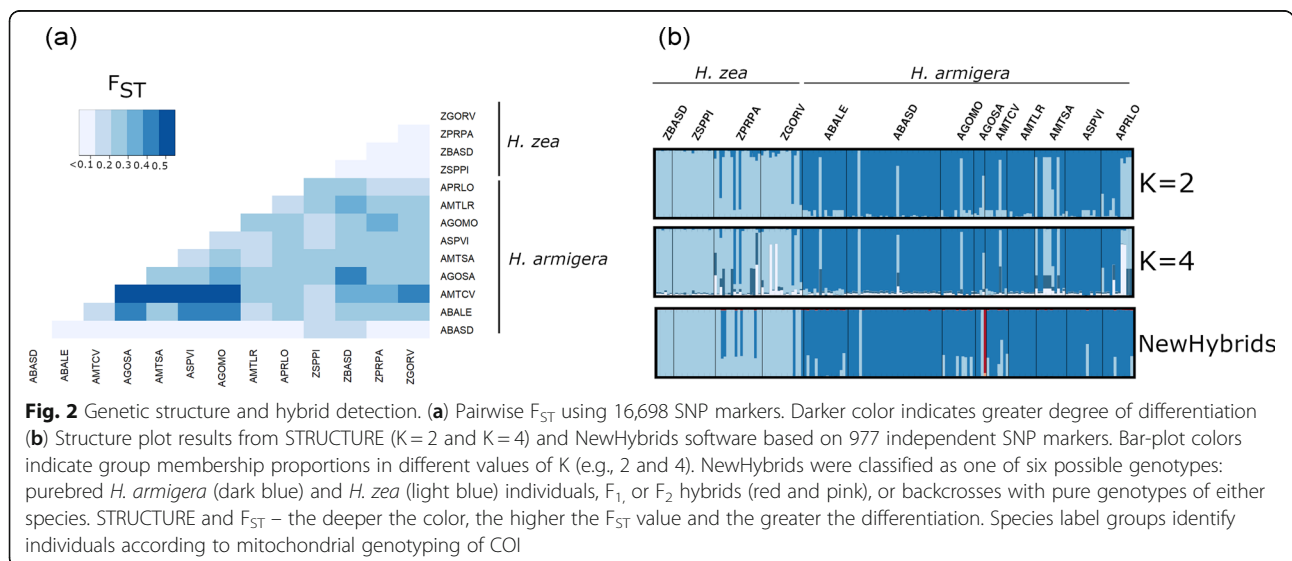


estimates among the *H. armigera* collection locations did not show a geographic pattern of structure (Fig. 2a).

The genetic divergence between *H. armigera* and *H. zea* samples can be clearly differentiated in the results from both STRUCTURE and NewHybrids; these analyses also showed a consistent presence of putative hybrids between the two species (Fig. 2b). Considering the information derived from the host plant, mitochondrial DNA, and SNPs to infer hybridization, our analyses concurred that 26 insects showed mixed ancestry (~15%), with an average mean rate of introgression of 10% and no significant differences between the species (*H. armigera*:  $\bar{x} = 0.15$ ,  $SD = 0.28$ ; *H. zea*:  $\bar{x} = 0.10$ ,  $SD = 0.25$ ) ( $\beta = -0.08$ ,  $SE = 0.06$ ,  $t$ -value =  $-1.35$ ,  $p < 0.18$ ). Bayesian assignment analyses indicated that the specimens of *H. zea* from two of

the four locations had pure ancestry (ZBASD and ZSPPI), whereas the specimens from the two remaining locations showed detectable levels of hybridization with *H. armigera* [ZPRPA:  $\hat{p} = 0.20$  (*armigera*);  $\hat{q} = 0.80$  (*zea*) and ZGORV:  $\hat{p} = 0.14$  (*armigera*);  $\hat{q} = 0.86$  (*zea*)] (Fig. 2b). A total of 7 out of 9 collection locations where mitochondrial DNA identified as *H. armigera* showed signs of hybridization, based on STRUCTURE and NewHybrids.

According to STRUCTURE, the most extensively “hybridized” location was PRLO [ $\hat{p} = 0.60$  (*armigera*);  $\hat{q} = 0.40$  (*zea*)], while NewHybrids identified AGOSA as the most extensively “hybridized” [ $\hat{p} = 0.50$  (*armigera*);  $\hat{q} = 0.50$  (*zea*)]. The NewHybrids approach detected fewer putative hybrids in our samples than structure analysis, and successfully flagged one putative F<sub>1</sub> hybrid in AGOSA (Fig. 2b, red vertical bar).



We used the STRUCTURE estimates of introgression ( $K = 2$ ), as our response variable used in subsequent linear mixed models. The results confirmed the significant effect of location, when species was the random factor ( $\beta = 0.29$ ,  $SE = 0.05$ ,  $t\text{-value} = 5.43$ ,  $p < 0.000$ ) (Fig. 3).

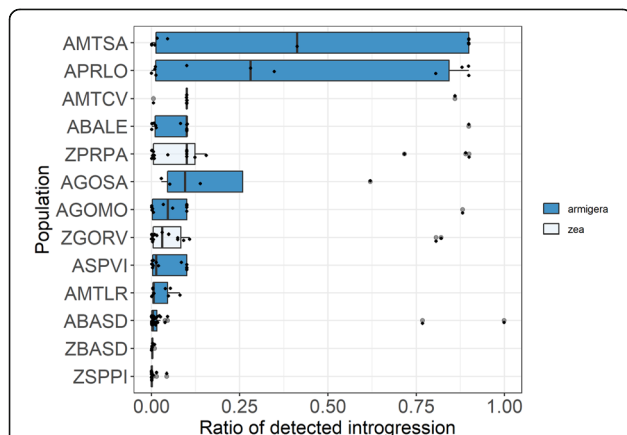
**Presence and direction of gene flow**

The results from Treemix largely agreed with the other inferences, successfully separating samples into two broad groups that corresponded to the long-term isolated lineages of *H. armigera* and *H. zea* (Fig. 4). Treemix also indicated at least three events of hybridization and one of admixture between *H. armigera* populations ( $m = 4$ ), based on the locations sampled. The main direction of interspecific gene flow seemed to be from *H. zea* to *H. armigera*, affecting insects from APRLO, AMCV, and AGOSA (Fig. 4).

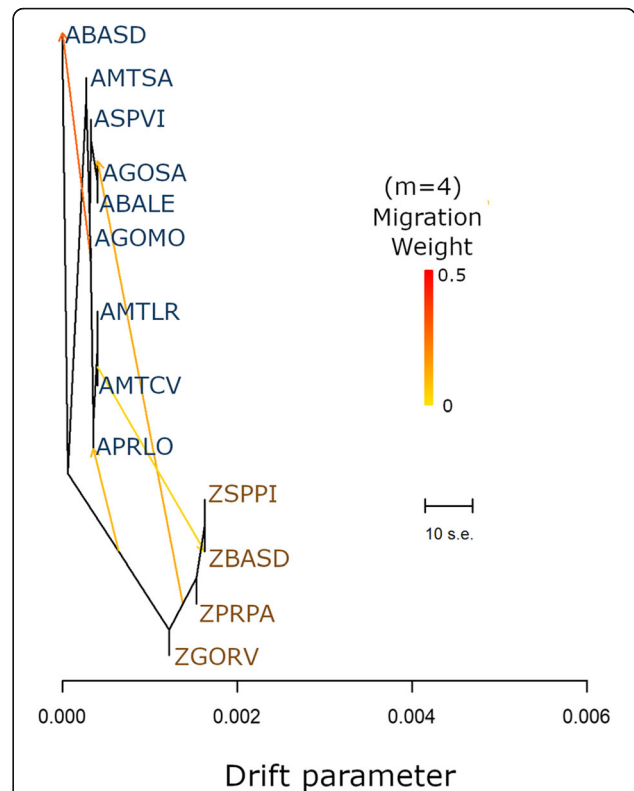
**Modeling the effects of landscape and environmental variables on hybridization**

The secondary contact between the two species was patchy and formed a mosaic of hybrid and non-hybrid zones (Fig. 5). To evaluate the potential impact of environmental variables on the rates of introgression between *H. zea* and *H. armigera* in Brazil, we orthogonally transformed climate and landscape variables, using two separate PCA analyses. The group of climate and landscape variables was first condensed into principal components, and the first axis of each PC was used as the predictor variable. The effects of population and species were controlled in our models.

Climate variables had significant effects on the introgression rate into *H. armigera* ( $\beta = 0.08$ ,  $SE = 0.03$ ,  $t\text{-value} = 3.49$ ,  $p < 0.00$ ) (Fig. 6a). The most important variable was the “mean temperature of the



**Fig. 3** Boxplot showing introgression proportions, using STRUCTURE ( $K = 2$ ) estimates across different geographical locations. Colors identify groups according to mitochondrial COI genotyping

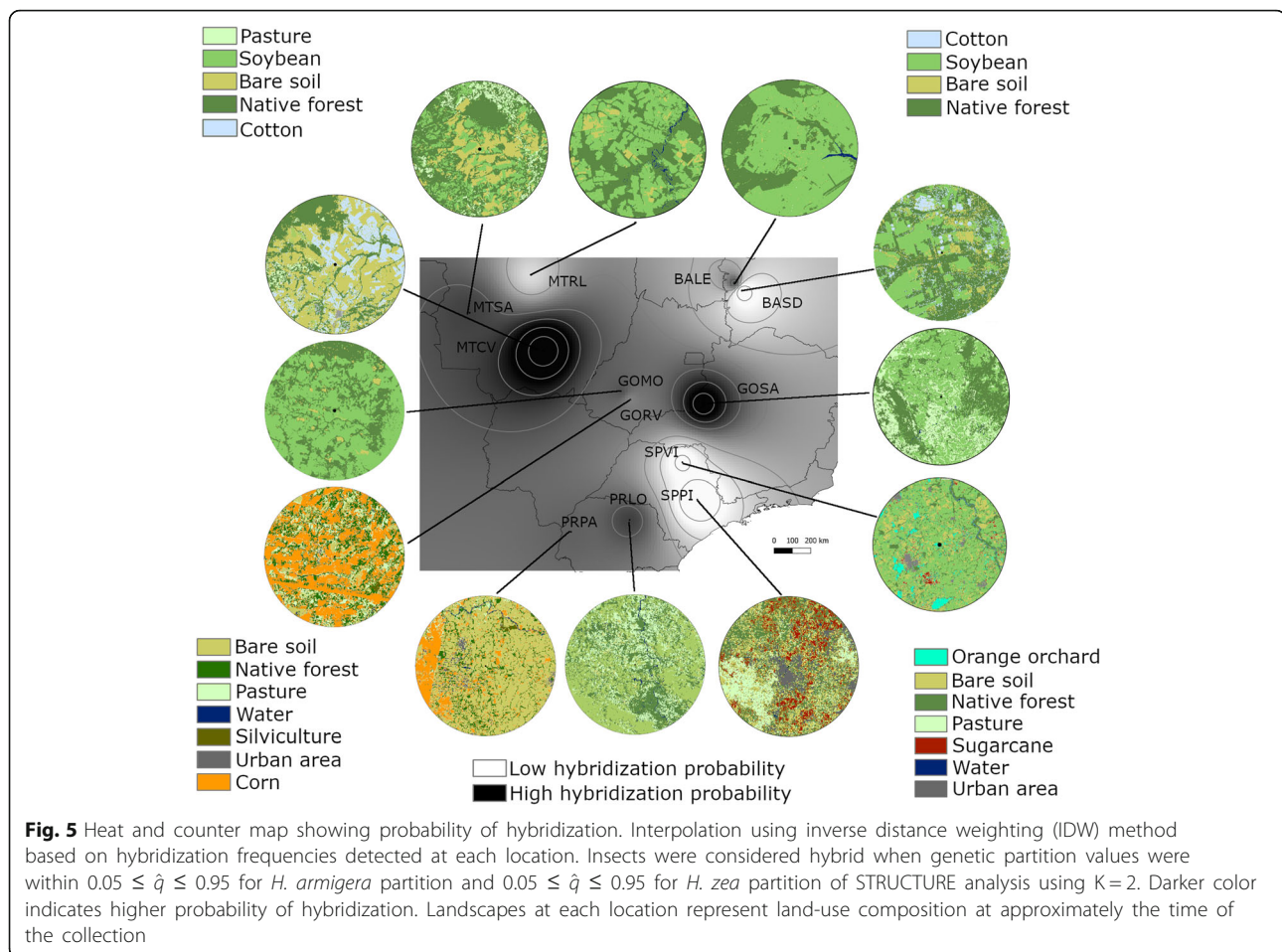


**Fig. 4** Maximum-likelihood tree constructed in Treemix based on 977 SNP markers with four migration events. Most migration events tended to move from *H. armigera* to *H. zea*. Among *H. armigera*, migration events occurred from AGOMO to ABASD

coldest quarter” (BIO11) (PC1 contribution = 7.96), “annual temperature range” (BIO1) (PC1 contribution = 7.46), and “precipitation in the driest month” (BIO14) (PC1 contribution = 7.16) on the first PC axis (58%). Evaluating the contribution of each location sampled to the first principal component, the largest variance contributions came from APRLO (PC1 contribution = 28.92) and APRPA (PC1 contribution = 23.21).

Landscape variables had a smaller but significant effect on the introgression rate in *H. armigera* ( $\beta = -0.09$ ,  $SE = 0.037$ ,  $t\text{-value} = -2.43$ ,  $p < 0.04$ ) (Fig. 6b). The cumulative variance on the first two axes contributed 58.14% of the total variance. The most important variable was maize (PC1 contribution = 22.41), followed by tree plantations (PC1 contribution = 13.94) and soybean (PC1 contribution = 13.86) on the first PC axis (32.8%). Evaluating the contribution of each sampling location to the first principal component, the largest variances came from ZPRPA (PC1 contribution = 42.3) and ASPPI (PC1 contribution = 25.01).

When we compared different model sets that contained explanatory variables (combined or individually), we



determined that the full model best explained the observed variance (AIC = 94.7), compared to the naïve model (AIC = 98.42,  $\chi^2 = 7.69$ ,  $p = 0.02$ ), climate-only model (AIC = 47.12,  $\chi^2 = 4.39$ ,  $p = 0.04$ ), or the landscape-only model (AIC = 100.05,  $\chi^2 = 7.32$ ,  $p = 0.006$ ).

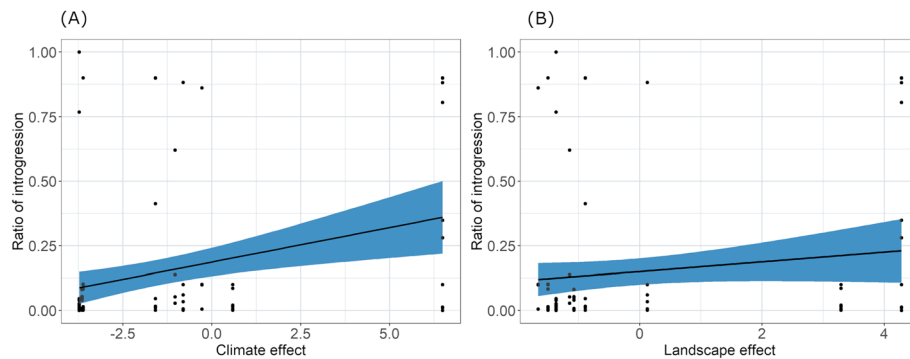
## Discussion

### Hybridization, asymmetric gene flow, and levels of introgression

Our data confirmed hybridization between *H. armigera* and *H. zea* in Brazilian crop fields [8, 9, 25]. Interspecific gene flow has occurred between *H. armigera* and its sibling taxon *H. zea* as a result of the secondary contact after 1.5 My of allopatric separation, and the consequences of this encounter are still unfolding [26, 27]. Previous research has established that hybridization between the two species was infrequent but possible under laboratory conditions [15, 16, 28], and more recent studies have collected evidence for hybridization in the field [8, 23, 25, 28]. However, the limitations of the genetic markers and the particular range of samples collected restricted interpretations within Brazil, which is the center for *H. armigera* invasion of the Americas. No sterility or

sex-ratio distortion has been observed in any previous study, but severe impairments in fitness are often reported, which can impact the practical viability of hybrids in the field [15, 28]. Here, we used thousands of genome-wide SNPs in tandem with secondary information including mtDNA markers, host-plant information, and morphological features to estimate hybridization between *H. armigera* and *H. zea* in different regions of Brazil. Our data showed that the hybridization varied significantly in the degree of introgression, depending on the sample location, landscape composition, and climate conditions.

We hypothesized that relatively few hybridization event occur but involve introgression of large genomic areas [8]. Our data supported this hypothesis with relative agreement between the different markers (SNPs and mitochondrial DNA), relatively small, but detectable levels of hybridization, and the high degree of compatibility and synteny of the two genomes. Despite laboratory evidence of hybrids' lack of fitness, continuous backcrossing in natural population can increase the compatibility of the introgressed material from the various recombinant types (i.e., "hybrid swarm") into the



**Fig. 6** Climate and landscape effects on introgression estimates. Introgression ratios were calculated based on STRUCTURE analysis using  $K = 2$ . Dependent variables were estimated from the first PC axis of the 19 Bioclim variables (climate variables) and 14 land-use classes (landscape variables)

pure lines, creating an adaptive bridge between the two species. Multiple hybridization events can enhance the fitness performance of the two species involved, even when the rate of hybridization is relatively limited [29]. Under these circumstances, hybridization allows adaptation to new climatic and landscape conditions encountered by the invading species [30].

Hybridization can also have profound repercussions for the native species, as demonstrated by the recent detection of the CYP337B3v2 resistance gene in *H. zea* [28]. This ubiquitous chimeric P450 gene confers pyrethroid resistance on *H. armigera* in Brazil [31], and now is present *H. zea*. This introgression provides compelling evidence for the potential adaptive advantage of hybridization, especially in agricultural systems. The implications can extend beyond insecticide resistance and affect other traits such as host range. For example, *H. zea* has lost a significant number of detoxification genes and gustatory receptors due to genetic drift, and might be ‘re-acquiring’ some of the ‘lost’ genes from *H. armigera* [26]. The scenario for hybridization in the Americas may become increasingly complex as *H. armigera* spreads and overlaps with another *Helicoverpa* pests in Argentina (*H. gelatopoeon*) and with *H. zea* populations from North America [25].

In relation to the viability of hybrids, the interaction between genetic and environmental factors has shaped and will continue to shape the distribution of genetic diversity, leading either to the development of *H. armigera* ecotypes or to fusion into a single panmictic population. However, in order to determine the evolutionary trajectory of the two genetic groups, hybrid viability must be determined. If hybrid viability proves to be limited, then maintenance of the two species is the most likely scenario. On the other hand, if hybrid fitness exceeds that of the purebred lines, then the species are expected to more closely resemble one another. Therefore, while hybrids may perform poorly in a laboratory setting, some

hybrids may have beneficial qualities that increase fitness in a complex mosaic of agroecosystem. Our data indicate that hybrids are present in natural populations; whether or not the level of hybridization will increase or not needs further investigation.

#### Genetic structure, founder event, and admixture

The three major features of the *H. armigera* invasion in South America are the high mitochondrial haplotype diversity (i.e., haplotypes shared with Asia, Africa, and Europe), the genetic similarity among distant parts of recently colonized areas, and the differences in regional dynamics (i.e., host availability and host composition) [8, 32, 33]. Our data showed a high degree of differentiation among some *H. armigera* sampling locations, which might suggest some level of genetic structure. In addition, the higher values of  $F_{ST}$  among *H. armigera* populations can be explained by hybridization with *H. zea*, the presence of multiple *H. armigera* lineages, genetic drift (i.e., bottlenecks), and differences in local dynamics (i.e., natural selection). Many questions remain regarding how the various invasion events occurred in Brazil, such as if the invasive specimens originated from a pool of founders of mixed ancestry or if the admixture occurred upon arrival. The patterns of genetic substructure and intra-species hybridization within *H. armigera* populations captured by our data may suggest the presence of multiple *H. armigera* lineages that are partially admixed. In contrast, the  $F_{ST}$  values for *H. zea* do not suggest a high degree of genetic differentiation, supporting the evidence for panmixia, at least within Brazil [25, 34, 35].

The genetic structure of *H. armigera* populations has always been a contentious topic of debate [36–38], where the most evident signs of population structure were only present at large geographical scales or when other lineages were taken into account [8, 23]. Populations of *H. armigera* seem to be experiencing extensive gene flow in other parts of the world [12]. We can, therefore, expect that the

genetic differences among populations in Brazil might decrease and stabilize over time. However, population genetic structure caused by geographical regions or season and crop variations have also been reported in different ecological contexts, suggesting that *H. armigera* may not reach a level of panmixia like *H. zea* [22, 38]. Continued efforts are needed to monitor *H. armigera*'s population structure, which will improve the predictions how resistance might spread.

#### **RAD-Seq for hybridization**

Similar to other previous research using SNP data, we have also detected a substantial rate of missing data caused by the interruption of the recognition site of restriction enzymes [23]. The high proportion of missing data may be evidence of a significant level of differentiation between species and within populations. Large amounts of missing data can create inconsistencies in quantifying introgression in natural populations and in estimates of genetic differentiation [23]. If criteria are too strict, SNP filtering will reduce the number of markers and select only highly conserved regions of the genome. In this case, biases in estimating the real introgression can suggest no or reduced hybridization. Alternatively, using a too-permissive filtering strategy may generate inconsistency in hybridization estimates, especially when multiple populations are compared, as the estimated diversity will mostly compare different regions of the genome. To overcome those difficulties, we included as many markers as possible while reducing the threshold for missing loci. While acknowledging that the RAD-seq protocol is prone to these forms of biases when distant groups are compared, the approach has reliably resolved datasets with higher rates of missing data (i.e., up to 90%) [8, 39]. Nonetheless, more research using whole-genome sequencing from insects collected in the field is necessary to confirm the values of introgression presented here.

#### **Environmental impact of climate and landscape on hybridization**

Rather than forming parallel clines where admixture can be easily recognized, *H. armigera* and *H. zea* hybridization instead formed a "mosaic hybrid zone" where the patchy hybridization hotspots have no apparent spatial pattern [40, 41]. A closer inspection indicates that the hybrid zones are habitat-dependent and mostly associated with maize and soybean production. In Brazil, *H. zea* is predominantly associated with maize, whereas *H. armigera* is often found on soybean and cotton. The mosaic configuration of the agricultural landscapes and the intensity of Brazilian farming (two to three crop seasons in a year) facilitate the simultaneous production or succession of suitable host types in the same area. Our

study provided evidence that rotation among crops can be particularly problematic and increase the probability of hybridization. Furthermore, both species have resistance to management practices: *H. armigera* is resistant to commonly used pyrethroids and *H. zea* is resistant to the *CryIAc Bt*-protein present in some transgenic soybean. Therefore, we can expect to see the first signs of insecticide or *Bt* resistance caused by introgressions and host-changing behavior in areas with extensive production of maize and soybean. In areas where these crops do not coexist, hybridization levels are likely extremely low, indicating that the appropriate choice of crops to rotate and the use of polycultures are essential for preventing and managing hybridization in *Helicoverpa*.

#### **Conclusions**

In summary, we have found strong evidence for hybridization between *H. armigera* and *H. zea* in Brazil. According to the different methods of inference, hybridization between the two species ranged from 15 to 30% among Brazilian locations. No significant asymmetry in hybridization between the two species was detected, but the probability of hybridization and the extent of the introgression were significantly affected by the environmental conditions, including climate and landscape composition. Insects from locations where maize and soybean were present tended to show high levels of hybridization. The most concerning finding is the continuous exchange of adaptive genetic variation that will likely affect the host range and insecticide resistance. If hybridization continues and increases it will likely complicate the management of these pests and further threaten crop production in Brazil. Continuous monitoring of the hybridization process is necessary because of the expansion of agricultural areas, climatic changes, the composition of crop species and varieties, and the planting seasons in South America. These constantly changing factors could lead to sudden changes in the rate of introgression between these *Helicoverpa* species, and strongly impact on the host range and resistance management.

#### **Methods**

##### **Sample collection, DNA extraction, and species identification**

Larvae of both *Helicoverpa* species were collected from 13 different Brazilian locations by active searching on host plants. The sampling included the most important soybean, cotton, and maize-producing areas in Brazil during the 2015 crop season. Detailed information about the host plant, collection date, and geographic coordinates is presented in Table 1. Upon collection, samples were preserved in pure ethanol and stored at  $-80^{\circ}\text{C}$  until further manipulation.

**Table 1** Information about sampled locations of 13 collection locations of *Helicoverpa* spp. in Brazil for SNP markers sequencing. N<sub>GEN</sub> refers to the number of insects successfully sequencing using SNP markers

Species	Locations	Host	Code	Date	Latitude	Longitude	N <sub>GEN</sub>
<i>H. armigera</i>	São Desidério, BA	Soybean, Cotton, Sorghum, Bean	ABASD	May 2015	12°26'27" S	45°26'47" W	34
<i>H. armigera</i>	L. E. Magalhaes, BA	Cotton	ABALE	June 2015	11°49'15" S	46°11'54" W	16
<i>H. armigera</i>	Campo Verde, MT	Cotton	AMTCV	May 2015	15°23'41" S	55°11'02" W	08
<i>H. armigera</i>	Lucas do Rio Verde, MT	Soybean	AMTLR	November 2014	11°40'57" S	55°47'49" W	10
<i>H. armigera</i>	Sapeza, MT	Cotton	AMTSA	May 2015	13°32'33" S	58°48'49" W	11
<i>H. armigera</i>	Montividiu, GO	Soybean	AGOMO	January 2015	17°22'30" S	51°23'33" W	12
<i>H. armigera</i>	S.A. do Rio Verde, GO	Soybean	AGOSA	February 2015	18°01'37" S	47°21'25" W	04
<i>H. armigera</i>	Londrina, PR	Soybean	APRLO	February 2015	23°41'46" S	50°57'52" W	11
<i>H. armigera</i>	Viradouro, SP	Soybean	ASPMI	January 2015	20°52'38" S	48°22'35" W	13
<i>H. zea</i>	São Desidério, BA	Maize	ZBASD	May 2015	12°26'27" S	45°26'47" W	6
<i>H. zea</i>	Rio Verde, GO	Maize	ZGORV	February 2015	17°28'03" S	51°07'43" W	15
<i>H. zea</i>	Palotina, PR	Maize	ZPRPA	February 2015	24°21'24" S	53°45'30" W	17
<i>H. zea</i>	Piracicaba, SP	Maize	ZSPPI	February 2015	22°41'50" S	47°38'34" W	15

Total DNA was extracted from each specimen following an adapted protocol based on the CTAB method [42]. After the DNA extraction, species identification was confirmed through a PCR-RFLP method involving the digestion of a mitochondrial fragment of the COI mitochondrial gene (~ 511 bp). The PCR reactions were prepared using the COI-F02/R02 set of primers, and the reaction product cut with the enzyme *BstZ17I* [43].

#### Genotyping by sequencing library preparation

A total of 172 samples of *Helicoverpa* species (53 *H. zea* and 119 *H. armigera*) were selected to generate two GBS libraries containing ~ 86 insects each [44]. Before the library-preparation step, the gDNA quality and quantity were assessed in each sample by visual inspection on agarose gel 1% (p/v), followed by determination of the concentration with a Qubit® 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA). We normalized DNA at 20 ng/μl and digested with a single restriction enzyme, endonuclease (*MseI*). Last, we used HiSeq 2500 to sequence the pair-end libraries, which were prepared and sequenced at the Molecular & Cellular Imaging Center Genomics Facility at the Ohio State University (Wooster, OH, USA). Raw fasta files of Illumina sequences were included in the SRA-NCBI repository (PRJNA615801).

#### Demultiplexing, SNP genotyping, and filtering strategy

Raw-sequence reads were demultiplexed using *process-radtags* implemented in STACKS 2.2 [45, 46]. Reads were trimmed at 85 bp after quality checking. In the first steps of the analysis, the program rescued RAD-tags from the reads, removed reads with uncalled bases, and then discarded reads with low-

quality scores (i.e., -r, -c, and -q). Several attempts were made to map the GBS reads to the reference genome (PRJNA378437); however, due to the low percentage of the alignment (< 15%), we decided to use the non-reference-based method available in STACKS. The *de-novo* approach to assemble loci has been extensively used in non-model system and when there is no reference genome available; this strategy is also more appropriate when the percentage of alignment is low. We ran the *de-novo* pipeline using all default parameters, closely following the method described by Anderson et al. [23]. After running preliminary tests, we concluded that the parameter combination used by Anderson et al. [23] provided the optimal yield regarding the number of markers retained and cluster resolution. Pair-end reads were integrated into a single-end locus, organized by loci in *tsv2bam* and assembled into contigs using *gstacks*. In the last step, we generated statistical summaries and Treemix analysis using the *population* module, allowing a minimum of 5% individuals required within groups and 100% between groups, excluding SNPs with less than 5% frequency, using one random SNP per RAD locus. Due to the great divergence between groups and the possibility of a high degree of variation within groups caused by hybridization, filtering parameters were relaxed, allowing an overall 20% presence of SNPs (i.e., to be included, a certain SNP must be shared with 20% of all individuals independently of their location). We conducted preliminary tests to maximize data retention while minimizing the rates of missing data in both species. The impact of hybridization varies in different parts of the genome, as previous studies have shown [47]. Thus, a different set of SNPs



isolated from different genome regions can potentially give different values of estimated introgression. Our approach will help identify and limit potential biases of the different imputation methods [48, 49].

### Nuclear admixture, introgression, and population structure

Species were identified using the collection information, including the host plant, morphological characters, and mtDNA genotyping, followed by the analysis with SNPs. Bayesian clustering methods implemented in STRUCTURE 2.3.4 and NewHybrids 1.1 were used to identify putative hybrids and to estimate proportions of nuclear admixture and patterns of introgression [50–52]. For parameter settings, we set the admixture model as the ancestry model and correlated frequencies as allele-frequency models. The posterior probability ( $q$ ), representing the proportion of the genotypes originating from cluster categories ( $K$ ), was later used to infer the putative degree of introgression in each sample. We used individual estimates of the introgression of insects collected at different locations as a dependent variable in models to explain possible causes of the observed differences.

First, we assumed  $K = 2$ , because two gene pools could potentially contribute to the genetic makeup of each sample. However, because strong evidence supports a history of multiple invasions of *H. armigera* [32], we also explored levels of substructure to detect the coexistence of different gene pools that may reflect the population structure of *H. armigera* in Brazil. We ran the STRUCTURE analysis for a range of  $K$  values ( $K = 1–10$ ), and subsequently used the Evanno method implemented in STRUCTURE HARVESTER 0.6.93 to test for the most likely number of  $K$  [53]. We used only one SNP per RAD locus (--write-random-snp) to minimize the effect of markers on linkage disequilibrium while performing long runs of the program to ensure convergence. We set the program to discard the first 150,000 steps (burn-in) and recorded 250,000 steps in each replicate ( $n = 10$ ). Replicates of each  $K$  value were aligned and averaged in CLUMPP 1.1.2 [54] and visualized in DISTRUCT 1.1 [55].

The number of clusters and the level of hybridization were also investigated using non-model-based methods such as Principal Components Analysis (PCA) with the R package *ade4* and *ade4*, as well as pairwise  $F_{ST}$  analysis [56, 57]. Additionally, we explored the phylogenetic relationships of insects collected at different locations, taking into account possible migration events, using the program Treemix [58]. Population divergence and migration events were estimated using bootstraps to calculate parameters in different scenarios by varying the number of migration events ( $m = 1–6$ ). The most likely

number of migration events was determined based on log-likelihood values and plotted residuals.

### Association studies with landscape and climatic variables

Given that populations of the two species are now in sympatry, interbreeding may occur at different rates, possibly related to the presences of their main agricultural hosts of soybean, cotton, and maize. To investigate the ecological context of hybridization of *H. armigera* and *H. zea* in Brazil, we considered two groups of environmental variables in our analysis: climatic and land-use variables. For the climatic variables, we used elevation and 19 locality-specific bioclimatic variables from the WorldClim database, with a resolution of 30 arc-seconds [59]. To account for the significant number of correlated inputs, a principal component analysis (PCA) was carried out to constrain the climatic variables, converting many climatic variables into a smaller set of linear, uncorrelated values. The linear models used climatic variables from the sampling location, using the first two PC coordinates, since they carry the most significant portion of the variance, while the importance of climatic variables was assessed based on their contribution to the PC axes.

Land-use (i.e. landscape) variables were obtained classifying agricultural-landscape components such as soybean, maize, and cotton. Landscapes also contained other crops such as sugarcane, tree plantations, and orange orchards, as well as non-crop elements such as native forests, pastures, water, and urban sites that were also included in the classification maps. We quantified and characterized landscape attributes, using satellite images with a maximum of 2 months of differences in the collected data. This time window was necessary due to image quality, cloudy weather, and the availability of satellite images in public databases. Two databases were used for the satellite image collections, the *Instituto Nacional de Pesquisas Espaciais* (INPE) and the United States Geological Survey – Earth Explorer (USGS/Earth Explorer), which provide images from CBERS 4 and LANDSAT 8, respectively.

We manipulated and classified the different attributes from satellite images using ArcGIS 10.2.2. Briefly, a buffer with a 25-km radius from the collection point was created to delimit the study area. This radius was chosen based on the relative size of regional agricultural areas, in order to prevent overlap between landscapes, and also based on the insect's flight capacity (up to 1000 km) [60]. Different signatures based on spectral responses can be linked to landscape attributes such as maize, soybean, and cotton, and therefore a supervised classification using the maximum-likelihood classification method was selected to separate classes within 25 km. The resulting classification was carefully revised and

manually curated to minimize classification errors, using information from crop calendars and by contacting growers in the respective areas. Similar to climate variables, we conducted PCA analyses of the standardized proportion of each class, using only the first two PCAs to generate the models.

We constructed linear mixed-effects models using the 'lmer' function in the R package 'lme4' to estimate the relative importance of environmental factors for *H. armigera* and *H. zea* hybridization in Brazilian croplands [61]. The average introgression rates for each population, estimated based on SNP data, were used as our response variable. We inspected the residuals of each variable for distortion in homoscedasticity, and normality by visually checking the diagnostic plot and the residual. The proportion of introgression was arcsine square root-transformed to correct for normality. First, we tested the effect of species (fixed effect = species), controlling for the effect of populations (random effect = populations) to assess the asymmetry in gene flow between the two species. Then, we used the first PCA coordinates as independent variables in a full model for the hybridization detected in *H. armigera*. We tested the effect of landscape and climate, using these variables as fixed factors while controlling for the effect of the population (random = populations). We checked the significance of the models by evaluating  $\chi^2$  and *p*-values from the likelihood-ratio test of model comparisons. The most complex models included the interaction between landscape and climate, followed by models of isolated factors, and naïve models.

#### Abbreviations

SNP: Single Nucleotide Polymorphism; CTAB: Cetrimonium bromide; DNA: Deoxyribonucleic acid; COI: Cytochrome c Oxidase Subunit I gene; PCR: Polymerase Chain Reaction; RFLP: Restriction Fragment Length Polymorphism; BstZ171: gene from *Bacillus stearothermophilus* 38 M (Z. Chen); GBS: Genotyping-by-sequencing; gDNA: Genomic Deoxyribonucleic acid; MSeI: gene from *Micrococcus* species (R. Morgan); RAD: Restriction site-associated DNA;  $F_{ST}$ : Fixation indices; mtDNA: Mitochondrial DNA; K: Cluster categories; PCA: Principal Component Analysis; PC: Principal Component; INPE: Instituto Nacional de Pesquisas Espaciais; USGS: United States Geological Survey; km: kilometer(s);  $\bar{x}$ : Mean; SD: Standard deviation; SE: Standard error;  $\beta$ : the slope of the regression line; *t*-value: Student's *t*-test value; *p*: probability value;  $\hat{q}$ : posterior mean estimates of fraction of the insect's genome from *H. armigera* ancestors;  $\hat{p}$ : Posterior mean estimates of fraction of the insect's genome from *H. zea* ancestors; AIC: Akaike information criterion

#### Acknowledgements

We thank SGS Gravena (SISBIO License #18018-1) and PROMIP (SISBIO License #40380-5) for helping to collect insect samples in different Brazilian regions. We are grateful to Guidolin A.S., of the Ohio State University, for her help in DNA purification. Research carried out using the computational resources of the Center for Mathematical Sciences Applied to Industry (CeMEAI) funded by FAPESP (grant 2013/07375-0).

#### Authors' contributions

APM, ASC, CO, EMGC and LMPG conceived the study. ARBN, JPB and LMPG collected the data. APM, ASC and CO provided reagents and sequencing financing. EMGC performed the analysis. ASC, EMGC, and

LMPG wrote the main manuscript. All authors contributed to writing and editing the manuscript. All authors read and approved the final manuscript.

#### Funding

We thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for the fellowship to EMGC (#2017/02393-0), fellowship to JPB (#2016/04159-2), fellowship to ARBN (#2016/09159-0 and #2014/26212-7), and financial support for the research to ASC (grant #2014/11495-3); the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (CAPES: Finance Code 001) for the Ph.D. scholarship to LMPG; and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial support for the research to CO (403851/2013-0) and fellowships for CO and ASC. The funding bodies had no role in the study design, data collection, analysis and interpretation, or drafting of the manuscript.

#### Availability of data and materials

The climatic data are available through the WorldClim Global Climate Database from the University of California, Berkeley. Raw fasta files of Illumina sequences were included in SRA-NCBI repository (PRJNA615801). We are providing a vcf file with the full SNP set along with files the list of markers used in the various analyses. See supplemental material available at Figshare: [https://figshare.com/articles/Hybridization\\_and\\_introgression\\_between\\_Helicoverpa\\_zea\\_and\\_H\\_armigera\\_an\\_adaptational\\_bridge/12201809](https://figshare.com/articles/Hybridization_and_introgression_between_Helicoverpa_zea_and_H_armigera_an_adaptational_bridge/12201809)

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Entomology and Acarology, University of São Paulo, Luiz de Queiroz College of Agriculture, Piracicaba, São Paulo 13418900, Brazil. <sup>2</sup>Department of Entomology & The Center for Applied Plant Sciences, Ohio Agricultural Research and Development Center, Thorne Hall, The Ohio State University, 1680 Madison Ave, Wooster, OH 44691, USA.

Received: 26 November 2019 Accepted: 29 April 2020

Published online: 25 May 2020

#### References

- Grgurinovica CA, Walsh D, Macbeth F. Eucalyptus rust caused by *Puccinia psidii* and the threat it poses to Australia. *EPPO Bull.* 2006;36:486–9.
- Eastale S. The history of introductions of *Bufo marinus* (Amphibia: Anura); a natural experiment in evolution. *Biol J Linn Soc.* 1981;16:95–113.
- Zayed A, Whitfield CW. A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proc Natl Acad Sci U S A.* 2008;105:3421–6.
- Ascunce MS, Yang CC, Oakey J, Calcaterra L, Wu WJ, Shih CJ, et al. Global invasion history of the fire ant *Solenopsis invicta*. *Science.* 2011;331:1066–8.
- Tay WT, Soria MF, Walsh T, Thomazoni D, Silvie P, Behere GT, et al. A brave New World for an Old World pest: *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Brazil. *PLoS One.* 2013;8:e80134.
- Floate K, Whitham T. The "hybrid bridge" hypothesis: host shifting via plant hybrid swarms. *Am Nat.* 1993;141:651–62.
- Pinheiro F, de Barros F, Palma-Silva C, Meyer D, Fay MF, Suzuki RM, et al. Hybridization and introgression across different ploidy levels in the Neotropical orchids *Epidendrum fulgens* and *E. puniceoluteum* (Orchidaceae). *Mol Ecol.* 2010;19:3981–94.
- Anderson CJ, Oakshott JGG, Tay WTT, Gordon KHJH, Zwick A, Walsh TKK. Hybridization and gene flow in the mega-pest lineage of moth, *Helicoverpa*. *Proc Natl Acad Sci USA.* 2018;115:5034–9.
- Mallet J. Invasive insect hybridizes with local pests. *Proc Natl Acad Sci U S A.* 2018;115:4819–21.
- Kriticos DJ, Ota N, Hutchison WD, Beddow J, Walsh T, Tay WT, et al. The potential distribution of invading *Helicoverpa armigera* in North America: is it just a matter of time? *PLoS One.* 2015;10:e0119618.

11. Yang Y, Li Y, Wu Y. Current status of insecticide resistance in *Helicoverpa armigera* after 15 years of Bt cotton planting in China. *J Econ Entomol.* 2013;106:375–81.
12. Jones CM, Parry H, Tay WT, Reynolds DR, Chapman JW. Movement ecology of pest *Helicoverpa*: implications for ongoing spread. *Annu Rev Entomol.* 2019;64:277–95.
13. Behere GT, Tay WT, Russell DA, Heckel DG, Appleton BR, Kranthi KR, et al. Mitochondrial DNA analysis of field populations of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and of its relationship to *H. zea*. *BMC Evol Biol.* 2007;7:117.
14. Mitter C, Poole RW, Matthews M. Biosystematics of the Heliiothinae (Lepidoptera: Noctuidae). *Annu Rev Entomol.* 1993;38:207–25.
15. Laster ML, Hardee DD. Inter-mating compatibility between north American *Helicoverpa zea* and *Heliiothis armigera* (Lepidoptera: Noctuidae) from Russia. *J Econ Entomol.* 1995;88:77–80.
16. Laster ML, Sheng CF. Search for hybrid sterility for *Helicoverpa zea* in crosses between the north American *H. zea* and *H. armigera* (Lepidoptera: Noctuidae) from China. *J Econ Entomol.* 1995;88:1288–91.
17. Daly JC, Gregg P. Genetic variation in *Heliiothis* in Australia: species identification and gene flow in the two pest species *H. armigera* (Hübner) and *H. punctigera* Wallengren (Lepidoptera: Noctuidae). *Bull Entomol Res.* 1985;75:169–84.
18. McKechnie SW, Spackman ME, Naughton NE, Kovacs IV, Ghosn M, Hoffmann AA. Assessing budworm population structure in Australia using the AT-rich region of mitochondrial DNA. *Proc Beltwide Cotton Insect Research Conference, New Orleans, vol. 2; 1993.* <http://agris.fao.org/agris-search/search.do?recordID=US9429452#Vz-x5aVbC8p.mendeley>.
19. Stokes NH, McKechnie SW, Forrester NW. Multiple allelic variation in a Sodium Channel gene from populations of Australian *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) detected via temperature gradient gel electrophoresis. *Aust J Entomol.* 1997;36:191–6.
20. Scott KD, Wilkinson KS, Lawrence N, Lange CL, Scott LJ, Merritt MA, et al. Gene-flow between populations of cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) is highly variable between years. *Bull Entomol Res.* 2005;95:381–92.
21. Ji YJ, Zhang DX, Hewitt GM, Kang L, Li DM. Polymorphic microsatellite loci for the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and some remarks on their isolation. *Mol Ecol Notes.* 2003;3:102–4.
22. Behere GT, Tay WT, Russell DA, Kranthi KR, Batterham P. Population genetic structure of the cotton bollworm *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) in India as inferred from EPIC-PCR DNA markers. *PLoS One.* 2013;8:e53448.
23. Anderson CJ, Tay WT, McGaughran A, Gordon K, Walsh TK. Population structure and gene flow in the global pest, *Helicoverpa armigera*. *Mol Ecol.* 2016;25:5296–311.
24. Silva CS, Cordeiro EMG, Paiva JB, Dourado PM, Carvalho RA, Head G, et al. Population expansion and genomic adaptation to agricultural environments of soybean looper, *Chrysodeixis includens*. *Evol Appl.* 2020. <https://doi.org/10.1111/eva.12966>.
25. Leite NA, Correa AS, Michel AP, Alves-Pereira A, Pavinato VAC, Zucchi MI, et al. Pan-American similarities in genetic structures of *Helicoverpa armigera* and *Helicoverpa zea* (Lepidoptera: Noctuidae) with implications for hybridization. *Environ Entomol.* 2017;46:1024–34.
26. Pearce SL, Clarke DF, East PD, Elfekih S, Gordon KHJ, Jermiin LS, et al. Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. *BMC Biol.* 2017;15:63.
27. Hardwick DF. The corn earworm complex. *Mem Entomol Soc Canada.* 2011; 97(S40):5–247.
28. Walsh TK, Jousen N, Tian K, McGaughran A, Anderson CJ, Qiu X, et al. Multiple recombination events between two cytochrome P450 loci contribute to global pyrethroid resistance in *Helicoverpa armigera*. *PLoS One.* 2018;13:e0197760.
29. Marescalchi O, Scali V. Automictic parthenogenesis in the diploid-triploid stick insect *Bacillus atticus* and its flexibility leading to heterospecific diploid hybrids. *Invertebr Reprod Dev.* 2003;43:163–72.
30. Lewontin RC, Birch LC. Hybridization as a source of variation for adaptation to new environments. *Evolution.* 2006;20:315–36.
31. Durigan MR, Corrêa AS, Pereira RM, Leite NA, Amado D, Sousa DR, et al. High frequency of CYP337B3 gene associated with control failures of *Helicoverpa armigera* with pyrethroid insecticides in Brazil. *Pestic Biochem Phys.* 2017;143:73–80.
32. Tay WT, Walsh TK, Downes S, Anderson C, Jermiin LS, Wong TKF, et al. Mitochondrial DNA and trade data support multiple origins of *Helicoverpa armigera* (Lepidoptera, Noctuidae) in Brazil. *Sci Rep.* 2017;7:45302.
33. Leite NA, Alves-Pereira A, Correa AS, Zucchi MI, Omoto C. Demographics and genetic variability of the New World bollworm (*Helicoverpa zea*) and the Old World bollworm (*Helicoverpa armigera*) in Brazil. *PLoS One.* 2014;9: e113286.
34. Seymour M, Perera OP, Fescemyer HW, Jackson RE, Fleischer SJ, Abel CA. Peripheral genetic structure of *Helicoverpa zea* indicates asymmetrical panmixia. *Ecol Evol.* 2016;6:3198–207.
35. Perera OP, Blanco CA. Microsatellite variation in *Helicoverpa zea* (Boddie) populations in the southern United States. *Southwest Entomol.* 2011;36: 271–87.
36. Endersby NM, Hoffmann AA, McKechnie SW, Weeks AR. Is there genetic structure in populations of *Helicoverpa armigera* from Australia? *Entomol Exp Appl.* 2007;122:253–63.
37. Scott KD, Lawrence N, Lange CL, Scott LJ, Wilkinson KS, Merritt MA, et al. Assessing moth migration and population structuring in *Helicoverpa armigera* (Lepidoptera: Noctuidae) at the regional scale: example from the Darling downs. *Australia J Econ Entomol.* 2005;98:2210–9.
38. Scott LJ, Lawrence N, Lange CL, Graham GC, Hardwick S, Rossiter L, et al. Population dynamics and gene flow of *Helicoverpa armigera* (Lepidoptera: Noctuidae) on cotton and grain crops in the Murrumbidgee Valley, Article. *J Econ Entomol.* 2006;99:155–63.
39. Tripp EA, Tsai YHE, Zhuang Y, Dexter KG. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecol Evol.* 2017;7:7920–36.
40. Rand DM, Harrison RG. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution.* 1989;43:432.
41. Ross CL, Harrison RG. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. *Evolution.* 2002;56:2296–312.
42. Azmat MA, Khan IA, Cheema HMN, Rajwana IA, Khan AS, Khan AA. Extraction of DNA suitable for PCR applications from mature leaves of *Mangifera indica* L. *J Zhejiang Univ Sci B.* 2012;13:239–43.
43. Behere GT, Tay WT, Russell DA, Batterham P. Molecular markers to discriminate among four pest species of *Helicoverpa* (Lepidoptera: Noctuidae). *Bull Entomol Res.* 2008;98:599–603.
44. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 2011;6:e19379.
45. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol.* 2013;22:3124–40.
46. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: Building and Genotyping Loci *De Novo* From Short-Read Sequences. *G3 – Genes Genom Genet.* 2011;1:171–82.
47. Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. *Mol Ecol.* 2016;25:2337–60.
48. Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. RAD seq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol Ecol.* 2013;22:3179–90.
49. Cariou M, Duret L, Charlat S. How and how much does RAD-seq bias genetic diversity estimates? *BMC Evol Biol.* 2016;16:240.
50. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155:945–59.
51. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics.* 2003;164:1567–87.
52. Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics.* 2002;160:1217–29.
53. Earl DA, von Holdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 2012;4:359–61.
54. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 2007;23:1801–6.
55. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes.* 2004;4:137–8.
56. Jombart T. *ade4*: an R package for the multivariate analysis of genetic markers. *Bioinformatics.* 2008;24:1403–5.

57. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* (NY). 1984;38:1358–70.
58. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012;8:e1002967.
59. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 2005;25:1965–78.
60. Pedgley DE, Tucker MR, Pawar CS. Windborne migration of *Heliothis armigera* (Hübner) (Lepidoptera: Noctuidae) in India. *Int J Trop Insect Sci*. 1987;8:599–604.
61. Bates D, Maechler M, Bolker B. lme4: Linear mixed-effects models using Eigen and Eigen. R Package version 1.1–7; 2013.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

